

A Machine Learning Approach for Social Media Content Filtering

Gayatri Jawharkar, Prof. Rahul M. Raut, Sarvesh Sonawane, Shubham Kandekar, Parag Thorat

Department of Information Technology
Sandip Institute of Technology and Research Centre, Nashik, India

Abstract: Communication has become stronger due to exponential increase in the usage of social media in the last few years. People use them for communicating with friends, finding new friends, updating any important activities of their life, etc. Due to their growing popularity and deep reach, these mediums are infiltrated with huge Volume of spam messages. Spam message randomly sent to multiple addressees by all sorts of groups, but mostly lazy advertisers and criminals who wish to lead you to phishing sites. In this we are using various machine learning techniques for detecting spam in the short text messages and also Google vision API for detecting spam images.

Keywords: Spam Filtration ,Google Vision API, OWASP, Naive bayes classifier, Dictionary Based Algorithm

REFERENCES

- [1] Gianluca et al. Stringhini. Detecting spammers on social networks. In Proceedings of the 26th Annual Computer Security Applications Conference, pages 1–9. ACM, 2010.
- [2] Chengfeng Lin et al. Analysis and identification of spamming behaviors in sina weibo microblog. In Proceedings of the 7th Workshop on Social Network Mining and Analysis. ACM, 2013.
- [3] Jong Myoung Kim, Zae Myung Kim, and Kwangjo Kim. An approach to spam comment detection through domain-independent features. In Big Data and Smart Computing (BigComp), 2016 International Conference on, pages 273–276. IEEE, 2016.
- [4] Chenwei Liu, Jiawei Wang, and Kai Lei. Detecting spam comments posted in micro-blogs using the self-extensible spam dictionary. In 2016 IEEE International Conference on Communications (ICC
- [5] Archana Bhattarai, Vasile Rus, and Dipankar Dasgupta. Characterizing comment spam in the blogosphere through content analysis. In Computational Intelligence in Cyber Security, 2009., pages 37–44. IEEE, 2009.
- [6] Michael Crawford, Taghi M. Khoshgoftaar, Joseph D. Prusa, Aaron N. Richter, and Hamzah Al Najada. Survey of review spam detection using machine learning techniques. In Journal of Big Data, volume 23, 2015.
- [7] Fangzhao Wu, Jinyun Shu, Yongfeng Huang, and Zhigang Yuan. Co-detecting social spammers and spam messages in microblogging via exploiting social contexts. In Neurocomputing, volume 201, pages 51–65, 2016.
- [8] T. M. Mahmoud and A. M. Mahfouz, “Sms spam filtering technique based on artificial immune system,” IJCSI International Journal of Computer Science Issues, vol. 9, no. 1, pp. 589–597, 2012.
- [9] X. Huang and M. Xu, “An Inter and Intra Transformer for Hate Speech Detection,” 2021 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST), 2021, pp. 346-349, doi: 10.1109/IAECST54258.2021.9695652
- [10] D. Sahnian, S. Dahiya, V. Goel, A. Bandhakavi and T. Chakraborty, “Better Prevent than React: Deep Stratified Learning to Predict Hate Intensity of Twitter Reply Chains,” 2021 IEEE International Conference on Data Mining (ICDM), 2021, pp. 549-558, doi: 10.1109/ICDM51629.2021.00066.
- [11] H. Watanabe, M. Bouazizi and T. Ohtsuki, “Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection,” in IEEE Access, vol. 6, pp. 13825-13835, 2018, doi: 10.1109/ACCESS.2018.2806394.

- [12] S. Alsafari and S. Sadaoui, "Semi-Supervised Self-Learning for Arabic Hate Speech Detection," 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2021, pp. 863-868, doi: 10.1109/SMC52423.2021.9659134.
- [13] K. -Y. Lin, R. K. -W. Lee, W. Gao and W. -C. Peng, "Early Prediction of Hate Speech Propagation," 2021 International Conference on Data Mining Workshops (ICDMW), 2021, pp. 967-974, doi: 10.1109/ICDMW53433.2021.00126.
- [14] R. A. Ilma, S. Hadi and A. Helen, "Twitter's Hate Speech Multi-label Classification Using Bidirectional Long Short-term Memory (BiLSTM) Method," 2021 International Conference on Artificial Intelligence and Big Data Analytics, 2021, pp. 93-99, doi: 10.1109/ICAIBDA53487.2021.